Point of confusion estimation using facial features and gaze tracking

Johan Locke
Georgia Institute of Technology
Atlanta, United States of America
jlocke33@gatech.edu

ABSTRACT
Confusion is an important part of the learning process, but prolonged periods of confusion can have a detrimental effect. Online education environments lack the benefit of face-to-face interaction and instantaneous detection of confusion by teachers. We built an inobtrusive, inexpensive and scalable confusion tracker using only a webcam and a web browser. In this first implementation, the confusion estimator outperforms randomly generated estimates and provides evidence of efficacy in support of an argument to conduct further research.

Author Keywords
confusion; gaze estimation; facial expression recognition; facial emotion recognition; computer vision; JavaScript;

ACM Classification Keywords
Applied computing

INTRODUCTION

The Impact of Confusion

Class sizes are growing due to the advancement of technology and the fast-paced adoption of online education in general. Digital and online education environments usually provide students with some autonomy and flexibility, but owing to its scale it is difficult to support students in a nuanced and personalized way [24,39].

Encountering difficulties while learning brings out an emotional response, most typically frustration and confusion. There is empirical evidence to show that if students are guided and supported during their periods of confusion, it can lead to productive learning outcomes [6,24]. Alternatively, if there is too much confusion for a prolonged period it leads to frustration and a lack of knowledge retention, which can lead to insurmountable learning difficulties later in life. [4,24,33]

"Confusion" is defined by the Farlex Partner Medical Dictionary as "A mental state in which reactions to environmental stimuli are inappropriate because the person is bewildered, perplexed, or unable to orientate herself or himself". Confusion is one of the main barriers to effective learning [27]. It influences a person's ability to concentrate, multitask, and can even create short term memory loss.

Seasoned teachers can detect confusion with relative ease, but confusion is a complex emotion and difficult to detect scientifically. Without face-to-face teaching environments, confusion is not only difficult to detect, but challenging to respond to [3,24].

Confusion is a difficult concept, or feeling, to isolate [5,27]. Confusion is most prevalent during complex-learning activities (tasks that require the student to draw inferences, answer causal questions, present coherent explanations and demonstrate a transfer of knowledge) [7]. Research has shown that confusion may be beneficial for conceptual learning, by allowing students to overcome misconceptions before they have a sophisticated understanding [24,33].

Detecting Confusion

Human emotions are observable as the result of the interactions between cognitive and affective functions [34]. Students and teachers must be able to monitor their progress and understand how to act based on their experience of difficulty or the reaching of an impasse [24].

Behavioral indicators of confusion are more predictive of learning than self-reports [7,33]. Previous studies have shown that students' interactions with an online learning system may be predictive of learning in the long-term, but imperfect in the short term. In isolation, frustration is a better predictor than confusion alone, but the measurement of both yields the most accurate results [31]. In essence, the pattern of students' transitions between emotional states are more predictive in terms of total learning, than each state on its own [33].

EMG, ECG and EEG Sensors

Typical sensors that can be successfully used to conduct emotion recognition analysis are electromyographs (EMGs), electrocardiograms (ECGs), electroencephalographs (EEGs), and cameras [16,20]. A sensor is an instrument that is used to detect or measure a physical phenomenon, and (1) may record information, or (2) may provide a response to the physical phenomenon. This could be as simple as a test subject's self-directed input or a complex technological device [22]. A confusion detection method is the physical behavior detected or measured by the researchers to collect the information. This could be a heart rate, a brain impulse signal or a click of a mouse on a screen to point to the area that confuses the student.

To measure knowledge retention, the Cognitive Absorption (CA) and Cognitive Load (CL) models are typically used. CA models how deeply immersive and enjoyable an experience is and is a requirement for lasting learning to

occur. It measures the psychophysiological mechanism during learning experiences using simulations. Research has shown that there exists a strong measurable psychophysiological link between multiple exogenous variables and the cognitive absorption in information technology students [4]. In contrast to the CA model, the CL model holds that learning acquisition is determined by the level of difficulty of the material, not the level of enjoyment. EEGs have been proven to be an effective tool for measuring CL.

Frustration is a better predictor of confusion than confusion alone. The measurement of both yields the most accurate results [31]. From a pedagogical point of view, exogenous-CL (the feeling of frustration when concepts are demonstrated poorly) needs to be minimized while germane-CL (the strain on short term memory) needs to be maximized for effective learning to occur [4].

EEGs, EMGs and ECGs are obtrusive in their nature. Compared to cameras, all the other sensors need to be worn by the test subject and as such places a limitation on the expense and scale at which a solution can be deployed and tested [33]. It is important to note that Conrad & Bliemel (2016) argue that recent advancements in the field of EEGs provide inexpensive wireless- and Bluetooth-enabled EEGs. This potentially allows for greater use in large scale experiments and broader consumer adoption.

Self-Report methods

Electronic "Muddy Cards" can be used to assist in self reports of confusion. Muddy cards capture pointed feedback from students during a lecture, like offline sticky notes or index cards. Students highlight the area of the video that is confusing to them, and lectures collate the feedback post-lecture to ascertain what part of the lecture needs to be adjusted or needs to receive greater attention in the next lecture. The use of online muddy cards have proven beneficial to both students and lecturers [9]. Like other self-reported measurements, they require conscious input by the student. In an ideal scenario, measurements should be subconscious and inobtrusive.

Apart from self-reports and the use of psychophysiological sensors, other studies have estimated confusion using student clicking behavior on course content [32,39], and by using sentiment analysis during forum participation [39].

D'Mello et al. (2014) measured confusion by using animated agents to present contradictory opinions, with students to decide which opinion carried merit, and proved that performance on multiple choice questions were higher when the contradictions were successful in confusing students.

Self-reporting of emotions is a problematic data collection method and usually interferes with the primary learning tasks. It lacks sensitivity due to social and cognitive biases and a prerequisite for participant is some level of emotional intelligence [3].

Facial Emotion Recognition (FER)

Face Recognition is the process of comparing the image of one or more known or unknown people to an existing database of people, and categorizing it as either verified/identified, or not [15,37]. It is a biometric approach that relies mainly on the person's physiological characteristics [37]. Facial Recognition can be extended using mathematical algorithms to recognize the person's expression and emotion [28].

Often in literature the acronym "FER" refers to either "Facial Expression Recognition" or "Facial Emotion Recognition", or both [16,28]. In this study we refer to FER as the recognition of emotional states based on facial expressions using mathematical algorithms to analyze faces in images or video.

Gaze Estimation and Gaze Tracking can determine what specifically a student is observing when they become confused. Gaze Estimation is the process of determining where a person is looking at a predefined ocular plane [2], and relies on three key concepts, namely Line of Gaze (LoG), Line of Sight (LoS) and Point of Regard (PoR) [8,14]. Gaze Tracking occurs when the ocular movements and Gaze Estimates are recorded over a period.

Under constrained conditions an inexpensive web camera can perform Facial Expression Recognition and Gaze Estimation close to commercial grade eye trackers. [30], and it can be performed in real time without the need of a Convolutional Neural Network (CNN) [17,30]. 2D FER using a standard web camera is possible, but it will be far less reliable. 3D-based FER methods outperform 2D approaches. By taking pose into account, the 3D approaches can better handle illumination variations [28].

There is empirical evidence that confusion can be identified, using a standard web camera, by looking for the activation of specific facial actions [3]

The efficiency of existing solutions

The ideal confusion sensor solution needs to be inexpensive, inobtrusive, scalable, be independent of content and context (environment), work with a wide demography of users, and most importantly, be reliable (see

Appendix – High-level Solution Requirements).

The degree to which each of the current solutions meet the requirements is listed in Table 1.

Sensor	Detection Method	Inexpensive	Inobtrusive	Scalable	Content	Environment	Demography	Reliable	Unweighted	Weighted
EMG	Electrical skeletal muscle activity	○	○	◔	●	○	●	◕	◔	◔
EEG	Electrical brain activity	◑	◔	◔	●	◑	●	◕	◑	◑
ECG	Electrical heart activity	○	○	◔	●	○	●	◕	◔	◔
Self-Report	Muddy cards	●	◕	●	◕	◕	◑	◑	◕	◑
	Clicking behavior	●	●	●	◑	◕	◑	◑	◕	◑
	Sentiment Analysis	●	●	●	◔	◕	◑	◑	◑	◑
	Choice selection	●	◕	●	◑	●	◑	◑	◕	◑
FER	2D without gaze estimation	◕	◕	●	●	●	◕	◔	◕	◑
	3D without gaze estimation	◑	●	●	●	●	◕	◑	◕	◕
	2D with gaze estimation	◕	◕	●	●	●	◕	◑	◕	◕
	3D with gaze estimation	◑	●	●	●	●	◕	◑	◕	◕

Table 1 - Solution efficiency at meeting requirements

Further notes regarding the measurement scale used:

- To calculate a weighted average, the weight of reliability was set to twice as important as any other requirement. Refer to "Appendix – Solution Requirement Weighting" for detail on solution requirement specific weights.
- The scale used to determine how well each sensor and detection method meets the requirements, is set out in "Appendix – Solution Requirement Measuring Scale".

EMG, EEG and ECG sensors were disqualified due to their obtrusive nature and inability to scale inexpensively.

Of the remaining sensors, seen in isolation, FER with Gaze Estimation meets the requirements the best when using a weighted and unweighted scale. This finding aligns well with Ko (2018) who noted that "a camera is the most promising type of sensor because it provides the most informative clues for FER [Facial Emotion Recognition] and does not need to be worn" and Huang et al. (2019) noted "Human facial image is the mainstream and promising input type, because it can provide abundant information for expression recognition research."

Both FER with Gaze Estimation and self-report sensors are inexpensive and a combination of the capability could potentially significantly increase reliability.

RELATED WORK

Emotions have been traditionally categorized according to two models, the Appraisal Theories (differentiation of emotions based on the relationship between the subject and the environment) and Dimensional Theories (representing a subject's emotions in a range of categories) [34]. Cowen and Keltner (2017) established that 27 distinct varieties of emotional experiences exit and prove that although emotions may be labelled discreetly, they are in fact continuous, which they calculate through a gradient of "meaning".

Numerous Artificial Intelligence (AI) software systems have been designed to mimic the process of human emotions [34]. These systems use Computational Models of Emotions (CMEs), which are designed to process emotional information, elicit synthetic emotions, and generate emotional behaviors. Today CMEs tend to be implemented constrained to a specific domain, with limited ability or opportunity to provide immediate feedback [34].

Learning Analytics is an emerging field which collects and analyses data that students produce while using a digital platform. The main tenant is that specific behaviors (such as mouse pointer movements, clicks, scrolls etc.) can quantify behaviors that align with differentiable emotional states [3].

Emotion AI, also known as "affective computing", allows for emotions to be detected, analyzed, and processed using voice and non-voice channels. Most importantly, it allows for the technology to provide a response. The voice signal is most used as input and it is suggested that by 2022 our virtual personal assistants will be more able to predict a person's emotional state than their family members [10,41].

There is significant potential benefit for humanity to be reaped from Emotion AI, including, but not limited to health care and adaptive learning, On the negative side, the technology has the potential to violate personal information privacy as detection can occur inobtrusively [26].

This study will implement the principles provided by Murdoch et al. (2020):

- Data should only be used with informed consent. The data collected should be aligned and constrained to the original objective.
- The target must be able to minimize harm and maximize value. There should be greater value to the data provider than the data collector.
- Models should be an aggregate of the population, rather than individual specific. Models should be dynamic and retrained constantly.
- Human consequence and agency, rather than AI, should be prioritized in shaping the outcome.

THE SOLUTION

There does not exist an inexpensive and inobtrusive solution that can remove the barrier that the online education environment creates between student and teacher for the identification of confusion

It has been shown that confusion can be a barrier to productive learning, and if unresolved can lead to long-term learning impairment and student disengagement. Boredom is one of the main contributors of disengagement in Australian Schools [11]. The problem is magnified as the world is moving towards online learning at an increasing pace, and teachers and developers of e-learning systems require tools that will assist in the monitoring progress and gaining an understanding of when and how to act based on when difficulty is experienced or an impasse is reached.

Existing proprietary technology exists to do both real-time Facial Emotion Recognition and Gaze Tracking, but its expense is prohibitive for wide-scale adoption. Notably, 2D web cameras are outperformed by their 3D counterparts under illumination and pose variations [28] and it becomes a question about what minimum level of accuracy must be obtained to be practically useful.

Three pertinent objectives was pursued whilst fulfilling all the ideal solution requirements:

- Identify the **moment of confusion**, based on a sequence of images from a standard 2D web camera, and Facial Emotion Recognition.
- Combine feedback from a standard 2D web camera-based gaze estimator and a facial emotion recognizer into a signal that identifies the **area of confusion**.
- Ensure **personal privacy**.

Objective 1 – The Moment of Confusion

It has been shown that open source capabilities exist to execute FER in real-time [29,30], but uncertainty remains as to how successful it will be at identifying micro-expressions. At an individual level, a "tell" is expected – like in poker – when a person is confused. What is unclear at this stage is whether that "tell" will be universal or individual.

Cowen and Keltner (2017) establishes that 27 distinct varieties of emotional experiences exits, using EEG. They prove that although emotions may be labelled discreetly, they are in fact continuous, which they calculate through a gradient of "meaning". It remains to be seen if the same conclusion can be drawn using FER. 2D FER using a standard web camera is possible, but it will be far less reliable [28].

Objective 2 – The Area of Confusion

Only a single light source is required to accurately determine the Point of Regard (POR) [35], and research has proven that an inexpensive 2D web camera can be used to estimate gaze accurately when the pose is constrained and the viewing plane is split into a small number of quadrants [17,23].

It is expected to be able to track the gaze at a "quadrant"-level, and to be able to convert the "area of confusion" into a heatmap.

It should be noted that the inclusion of a Convolutional Neural Network (CNN) does improve the precision of the Gaze Estimation [21], and should be considered for inclusion where the architecture allows for it.

Objective 3 – Ensure Personal Privacy

The system will need to adhere to the five fundamental principles (notice, choice, access, security and enforcement) of information privacy protection as defined by the Federal Trade Commission [13]. It is expected that each of those requirements will be met, except for enforcement which is governed by the students' government.

Since students will have to provide very personal information – their "micro-emotions" – for processing, it is important consider personal privacy to ensure future use of any technology built. Privacy assurances (also called "opt-in" in other literature) do not provide a direct or moderating effect on information disclosure [1]. User fear will need to be tempered ahead of, and throughout, the monitoring process.

A person's intention and willingness to disclose personal information is based on the risk-benefit calculus of doing so [1]. Internet users have six types of behavioral responses (information privacy-protective responses) when privacy concerns arise (refusal, misrepresentation, removal, negative, word-of-mouth, complaining directly to online companies, and complaining indirectly to third-party organizations) [36]. These concerns would have to be dealt with to ensure participation in a new technology.

Technology exists that will allow for the collection of Gaze Estimation and FER to be computed on the client-side, with no need to send personal information to a service provider, other than a probability of the end user being confused or not [25,29,30].

Architecture

The solution has several key components (see Figure 1):

- A gaze estimator
- A facial expression/emotion recognizer
- A moment of confusion identifier
- An area of confusion isolator

The high-level architecture (see "Appendix – Architecture") followed a classic web-based architecture including a (1) web browser, (2) application & web server, and (3) database server.

- A pure HTML5 and JavaScript implementation was chosen which required no additional software installation, thereby not incurring any additional end-user costs. For more information of the architecture, refer to "Appendix – Architecture".

Process flow

The platform leads the end-user through several stages:

1. Determine privacy constraints
2. Capture demographic information
3. Calibrate gaze tracking
4. Record facial expressions and gaze estimates as several questions are presented to the respondent
5. Estimate the area of confusion and report the results to the user
6. Capture respondent's self-reported area of confusion and their rating on the system's estimate
7. Record experiment results in a database for future development of a CNN (Convolutional Neural Network)

Some of these stages, like calibrating the gaze estimator, will not be explicitly visible to the user - this ensures the solution to remain inobtrusive. The end goal is to accurately predict a user's area of confusion, after having established the moment of confusion.

METHODOLOGY

Five (5) original abstract artworks (see "Appendix – Images of Abstract paintings used during experimentation") were shown to respondents. Each artwork was initially rotated by 90◦ from its original orientation. Images were cropped to remove artist names and to fit a perfectly square resolution of 500x500 pixels.

Respondents were asked to rotate the artwork into what they believed was the artwork's original orientation, clicking with their mouse anywhere on the artwork. Gaze estimates, facial expressions, and rotation events (mouse clicks) were recorded throughout the process. User could move onto the next artwork whenever they saw fit but could not navigate backwards to a previous artwork once done.

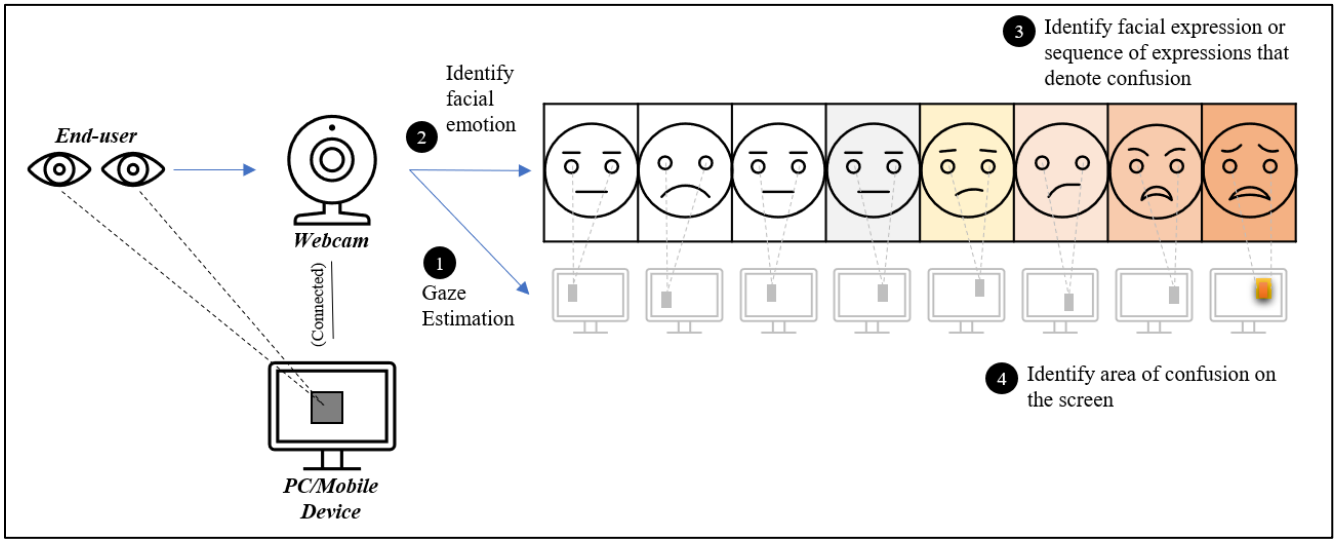


Figure 1 - Key solution components

Capturing of Estimates

On average, gaze estimates were more frequently calculable than expression estimates. Gaze estimations (G) were recorded independently (asynchronously and in parrallel) of facial expression estimates (E). To correlate gaze and expression estimates, gaze estimates required an adjustment calculation to align to the point in time (t) of the facial expression estimate.

Because gaze was being captured against the image of an artwork that had been rotated and that could be further rotated by the user, gaze estimations had to be normalised to the same 0° (zero degree) orientation, as can be seen in Figure 2.

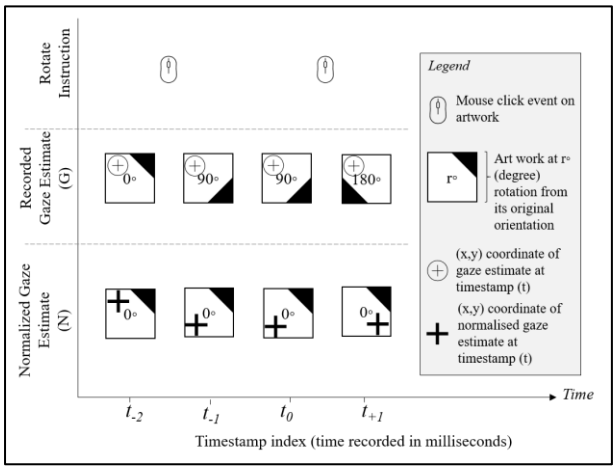


Figure 2 - Normalizing recorded gaze estimations over time

The normalized gaze estimation coordinates (N), were calculated as follows:

$$N_{t,(x',y')} = \boldsymbol{unrotate}(\boldsymbol{G_{t,r,(x,y)}})$$

Where:

- $[(t, r, (x, y))]$ is the recorded gaze estimate at coordinate position (x, y) with the artwork rotated by $r°$ (degrees), at time t.
- (x', y') are the normalised coordinates at time t.
- The $unrotate$ function rotates the coordinates around the center axis of the image back to a zero-degree orientation.

Facial expression estimates were recorded as the probability of a specific expression being observed at time t. Taking F being the complete set of facial expressions and

$$f \in F = \begin{Bmatrix} neutral, happy, sad, angry, \\ fearful, disgusted, surprised \end{Bmatrix}$$

Then E is the recorded facial expression at time t, or $E_t(f)$.

All expression probabilities at a specific point in time t sum to one.

$$\sum_{\forall f \in F} E_t(f) = 1$$

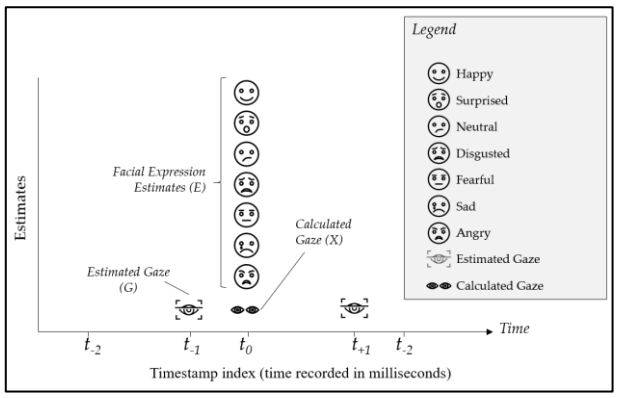


Figure 3 - Calculating the gaze estimate at the time of facial expression estimate

To calculate the gaze position at the timestamp of the facial expression estimate (X), the euclidean distance between the gaze position coordinate before the facial expression, and the next gaze position coordinate after the facial expression estimate, was calculated by weighting their relative time difference from the facial expression estimate, as can be seen in Figure 3.

l is the time lapsed between t_{+1} and t_{-1}, and $w_{t_{-1}}$ and $w_{t_{+1}}$ is the weighting of the time portion observed ahead and after E_t respectively.

$$l = t_{+1} - t_{-1}$$

$$w_{t_{-1}} = 1 - \frac{t_0 - t_{-1}}{l}, w_{t_{+1}} = 1 - \frac{t_{+1} - t_0}{l}$$

The calculated-gaze coordinate $X_{t_0} = (x''_{t_0}, y''_{t_0})$ at time t_0 is calculated as follows:

$$x''_{t_0} = x'_{t_{-1}} w_{t_{-1}} + x'_{t_{+1}} w_{t_{+1}}$$

$$y''_{t_0} = y'_{t_{-1}} w_{t_{-1}} + y'_{t_{+1}} w_{t_{+1}}$$

Estimating the area of confusion

A simplistic model was used. Only negative expressions (N) were used to calculate the area of confusion. With $n \in N = \{sad, angry, fearful, disgusted\}$, we assigned a weight value ($C_{t,(x'',y'')}$) to the calculated-gaze coordinate at time t.

$$C_{t,(x'',y'')} = \sum_{\forall n \in N} E_t(n)$$

The estimated point of confusion was calculated as the weighted average (x'', y''), each point weighted by their respective $C_{t,(x'',y'')}$.

Test population

The solution was hosted on a publicly accessible website (Microsoft Azure) and visitors from a reading-age could participate. 42 participants were sourced from 3 different groups, as detailed in "Appendix – Test Population".

The actual rate of participation was very low. As emphasis was placed on providing respondents a personal privacy briefing ahead of the experiment, 35.71% of respondents chose not to proceed through to the actual confusion tracking experiment, as can be seen in Figure 4.

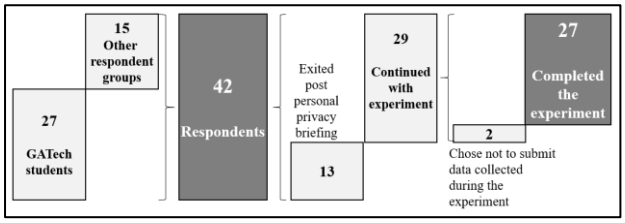


Figure 4 - Experiment respondent completion paths

To reduce negative data harvesting perceptions which may be associated with a software application of this kind, very little demographic information (age, gender) was collected directly, with the software estimating the self-reported values for comparative reasons. Summary population statistics of those 27 who completed the experiment, are shown in Table 2.

Variable	Statistic	Self-Reported	AI Estimated
Age (years)	Minimum	6	6
	Maximum	56	61
Female	Count	11	9
Male	Count	16	18

Table 2 - Summary population statistics

THE RESULTS

Area of Confusion Accuracy

Users were presented with a heatmap of estimated areas of confusion and reported their perceived accuracy of the confusion tracker using a five-point rating scale (1 = Very Bad, 5 = Perfect), as well as the x and y coordinate of the point the deemed confused them the most about the orientation of the artwork. See Table 3 for the summary results of the aggregated self-reported area of confusion per artwork.

Artwork #	Mean User Rating Estimator Correctness	% Respondents selecting correct final orientation
1	3.8	82%
2	3.8	57%
3	3.4	50%
4	3.6	11%
5	3.3	61%
Overall	**3.6**	**52%**

Table 3 - Mean respondent rating of the accuracy of the area of confusion

Post-experiment the population average self-reported and estimated area of confusion were calculated, as shown in Table 4. The error is calculated as the average Euclidean distance error in pixels.

Artwork #	Confusion (x,y) coordinate									
	Self-Reported				Estimated				Error	
	Mean		Std Dev		Mean		Std Dev		Mean	Std Dev
	x	y	x	y	x	y	x	y		
1	266	219	111	116	282	286	83	77	185	83
2	210	253	108	64	215	279	76	97	134	51
3	252	181	110	115	187	264	88	76	198	84
4	230	216	99	99	181	232	69	70	157	64
5	261	256	100	126	215	267	98	73	177	99
Population average									**170**	**80**
Self-reported compared against 1000 randomly generated estimated coordinates									232	93
Both self-reported and estimated coordinates randomly generated, average (Refer "Appendix – Random location generator")									261	124

Table 4 - Population average Self-Reported and Estmated Area of Confusion (measurements is pixels)

The confusion estimator outperformed randomly generated estimates. At this stage of the research, the magnitude of improvement that the estimator offer is not important - it provides evidence of efficacy and support for an argument to conduct further research. The aggregated average self-reported and estimated results for each artwork shown in Figure 5.

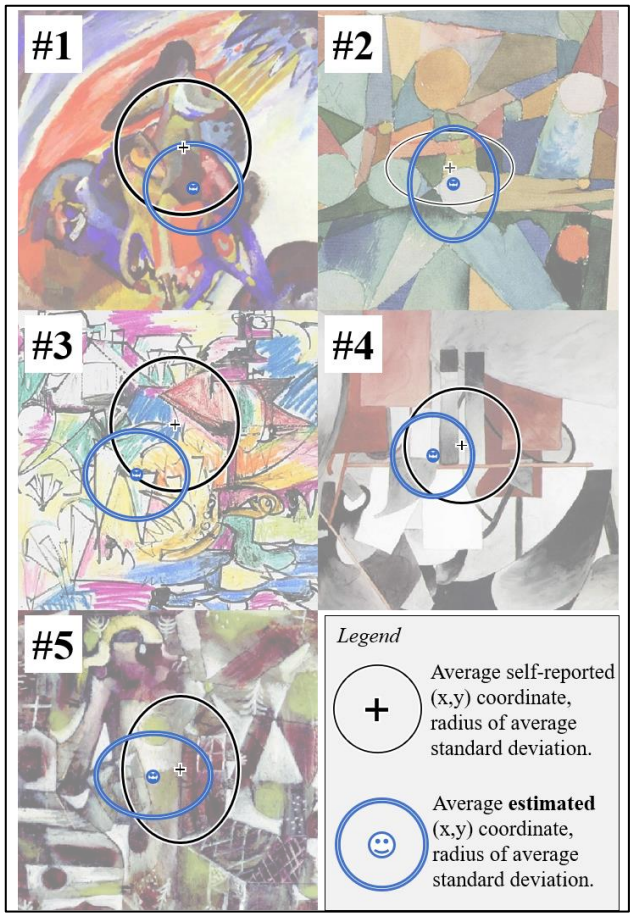


Figure 5 - Average (x,y) coordinate for self-reported and estimated point of confusion

Using the artwork experiment, respondents were only able to select the correct orientation of the artwork 52% of the time (see Table 3). We believe we were successfully able to induce confusion.

Ease and fear of use

27 current Georgia Institute of Technology CS 6460 (Educational Technology) students completed the experiment and were asked to also complete an additional short survey post the experiment. The system was rated at 4.1 out of 5 on ease of use. On the question "How comfortable were you to share your facial feature data?", the average response was 3.5 out of 5.

Performance

Confusion is a micro-expression and is usually accompanied by frustration. Even basic emotions like anger and fear generally have the lowest recognition rates [28]. Micro expressions are spontaneous and subtle facial movements that occur involuntarily. They reveal the true and potential expressions of a person for only a limited time (less than 1/3 s) [16].

To capture this, the system would have to reliably measure gaze and facial expression faster than once every 333 ms (milliseconds), or at rate of more than 3 fps (frames per second). As can be seen in Table 5, this was achieved on average, and notably without optimization of any code interaction between the two estimators.

Estimation Frequency	Minimum	Average	Max
Gaze	31 ms	93 ms	188 ms
	32.75 fps	10.69 fps	5.31 fps
Facial	129ms	328 ms	1453 ms
Expression	7.7 fps	3.04 fps	0.69 fps

Table 5 - Estimation frequency Results

LIMITATIONS

As noted earlier, the current implementation of the area of confusion estimator is simplistic and crude. Apart from its limitations and requirement for further development, the following should be noted:

- The solution utilized webgazer.js gaze estimator, which is known to be susceptible to pose changes. To counteract this, (1) users were requested to constrain their pose as much as possible during the assessment, and (2) the gaze estimator was continuously re-calibrated in the background.
- The gaze and facial expression estimators worked independently, which sacrificed computational performance. However, in the initial implementation it was deemed more important to use off-the-shelf open source capabilities than to adapt them for performance reasons.
- Mobile devices, such as iPads and iPhones were able to load the experiment, but as the solution was built around mouse clicks and movements, it yielded very poor results. Users were actively discouraged to use these devices during testing and to use a desktop computing device or a laptop.
- Because the webgaser.js library relies on mouse movements and click events, it does provide interference to the estimated gaze location.
- Respondents with more than one web camera could not select which one the system should utilize. One respondent also noted that the system did not function correctly when the web cam was located at the bottom of the screen.
- Several participants noted that it is not clear, even to them, what they found most confusing about the orientation of the artwork. As noted in the earlier sections of this paper, confusion is an elusive concept.

CONCLUSION

The solution outperformed randomly generated estimates, but the current incarnation of the area of confusion estimator is far too crude to lead to any formal conclusion based on this small number of respondents.

The experiment used abstract art works to detect confusion by asking respondent to only select one single point of confusion. This clearly does not capture a complete view of what was most confusing at an overall level. Consequently, comparing this self-reported point to a single average estimated point of confusion is potentially similarly flawed.

This research has shown that respondents retained concern about their personal privacy post viewing the personal privacy slides. Even though the solution appears to be easy to use, the solution has a long way to go to prove to end-users that it offers more value to them by choosing to use it than not to use it.

The system was performant on average, but for half of the population the average frequency of facial expression estimations were too slow. The solution does have room for optimization, but it is unclear at this stage how much impact it would have.

FUTURE WORK

One of the objectives of the solution was to identify the moment of confusion. This could not be completed during this initial experiment.

Facial Action Units (AU), specifically AU4 (lowering of the eyebrows), is said to be a far more accurate predictor of confusion that facial expressions alone [12]. Facial feature locations have been recorded during this experiment and could be used to build an estimator using this data.

In future studies, a control group should be utilized that employs a confusion sensor (such as an EEG) to compare experiment results.

The inclusion of a Convolutional Neural Network (CNN) does improve the precision of the Gaze Estimation [21], and should be considered for inclusion where the architecture allows for it

REFERENCES

1. Ibrahim M. Al-Jabri, Mustafa I. Eid, and Amer Abed. 2019. The willingness to disclose personal information: Trade-off between privacy concerns and benefits. *Information & Computer Security* ahead-of-print, ahead-of-print.

2. F. Alnajar, T. Gevers, R. Valenti, and S. Ghebreab. 2013. Calibration-Free Gaze Estimation Using Human Gaze Patterns. *2013 IEEE International Conference on Computer Vision*, 137–144.

3. Amaël Arguel, Lori Lockyer, Ottmar V. Lipp, Jason M. Lodge, and Gregor Kennedy. 2017. Inside Out: Detecting Learners' Confusion to Improve Interactive Digital Learning Environments. *Journal of Educational Computing Research* 55, 4: 526–551.

4. Colin Conrad and Michael Bliemel. 2016. Psychophysiological Measures of Cognitive Absorption and Cognitive Load in E-Learning Applications. *ICIS 2016 Proceedings*.

5. Alan S. Cowen and Dacher Keltner. 2017. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*.

6. Linda Darling-Hammond, Lisa Flook, Channa Cook-Harvey, Brigid Barron, and David Osher. 2020. Implications for educational practice of the science of learning and development. *Applied Developmental Science* 24, 2: 97–140.

7. Sidney D'Mello, Blair Lehman, Reinhard Pekrun, and Art Graesser. 2014. Confusion can be beneficial for learning. *Learning and Instruction* 29: 153–170.

8. Reza Ghiass and Denis Laurendeau. 2018. Highly Accurate and Fully Automatic 3D Head Pose Estimation and Eye Gaze Estimation Using RGB-3D Sensors and 3D Morphable Models. *Sensors* 18: 4280.

9. Elena L. Glassman, Juho Kim, Andrés Monroy-Hernández, and Meredith Ringel Morris. 2015. Mudslide: A Spatially Anchored Census of Student Confusion for Online Lecture Videos. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Association for Computing Machinery, 1555–1564.

10. Laurence Goasduff. 2018. Emotion AI Will Personalize Interactions. *Emotion AI Will Personalize Interactions*. Retrieved June 14, 2020 from //www.gartner.com/smarterwithgartner/emotion-ai-will-personalize-interactions/.

11. Peter Goss, Juli Sonnemenn, and Grattan Institute. 2017. *Engaging students : creating classrooms that improve learning*. Carlton : Grattan Institute.

12. Joseph F. Grafsgaard, Kristy Elizabeth Boyer, and James C. Lester. 2011. Predicting Facial Indicators of Confusion with Hidden Markov Models. *ACII*.

13. Robert W. Hahn and Anne Layne-Farrar. 2002. The Benefits and Costs of Online Privacy Legislation. *Administrative Law Review* 54, 1: 85–172.

14. Dan Hansen and Qiang Ji. 2010. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE transactions on pattern analysis and machine intelligence* 32: 478–500.

15. Nawaf Hazim Barnouti, Sinan Sameer Mahmood Al-Dabbagh, and Wael Esam Matti. 2016. Face Recognition: A Literature Review. *International Journal of Applied Information Systems* 11, 4: 21–31.

16. Yunxin Huang, Fei Chen, Shaohe Lv, and Xiaodong Wang. 2019. Facial Expression Recognition: A Survey. *Symmetry* 11, 10: 1189.

17. Shubham Jain and Enda Fallon. 2019. Low-cost Gaze Detection with Real-time Ocular Movements Using Coordinate- Convolutional Neural Networks. *International journal of simulation: systems, science & technology*.

18. JS Foundation- js.foundation. jQuery. Retrieved July 15, 2020 from https://jquery.com/.

19. Nikhil Karkra. 2020. Face recognition using JavaScript - DEV. Retrieved July 15, 2020 from https://dev.to/karkranikhil/face-recognition-using-javascript-33n5.

20. Byoung Chul Ko. 2018. A Brief Review of Facial Emotion Recognition Based on Visual Information. *Sensors* 18, 2: 401.

21. Kyle Krafka, Aditya Khosla, Petr Kellnhofer, et al. 2016. Eye Tracking for Everyone. .

22. Bibeg Hang Limbu, Halszka Jarodzka, Roland Klemke, and Marcus Specht. 2018. Using sensors and augmented reality to train apprentices using recorded expert performance: A systematic literature review. *Educational Research Review* 25: 1–22.

23. Yi Liu, Bu Lee, Deepu Rajan, Andrzej Sluzek, and Martin McKeown. 2019. CamType: assistive text entry using gaze with an off-the-shelf webcam. *Machine Vision and Applications*.

24. Jason M. Lodge, Gregor Kennedy, Lori Lockyer, Amael Arguel, and Mariya Pachman. 2018. Understanding Difficulties and Resulting Confusion in Learning: An Integrative Review. *Frontiers in Education* 3.

25. Vincent Mühler. 2020. *justadudewhohacks/face-api.js*. .

26. Robin Murdoch, Agneta Björnsjö, Paul Johnson, and Marc Flynn. 2020. Getting Emotional: How Platforms, Technology and Communications companies can build a responsible future for Emotional AI. *Accenture Research*: 24.

27. Zhaoheng Ni, Ahmet Cem Yuksel, Xiuyan Ni, Michael I Mandel, and Lei Xie. 2017. Confused or not Confused? *ACM-BCB ... ...: the ... ACM Conference on Bioinformatics, Computational Biology and Biomedicine. ACM Conference on Bioinformatics, Computational Biology and Biomedicine* 2017: 241–246.

28. Francesca Nonis, Nicole Dagnes, Federica Marcolin, and Enrico Vezzetti. 2019. 3D Approaches and Challenges in Facial Expression Recognition Algorithms—A Literature Review. *Applied Sciences* 9, 18: 3904.

29. Alexandra Papoutsaki, James Laskey, and Jeff Huang. 2017. SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search. *Proceedings of the ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR)*, ACM.

30. Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediyana Daskalova, Jeff Huang, and James Hays. 2016. WebGazer: Scalable Webcam Eye Tracking Using User Interactions. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, AAAI, 3839–3845.

31. Zhongxiu Peddycord-Liu, Visit Pataranutaporn, Jaclyn Ocumpaugh, and Ryan Baker. 2013. Sequences of Frustration and Confusion, and Learning. .

32. J. Elizabeth Richey, Juan Miguel L. Andres-Bray, Michael Mogessie, et al. 2019. More confusion and frustration, better learning: The impact of erroneous examples. *Computers & Education* 139: 173–190.

33. J. Elizabeth Richey, Bruce M. McLaren, Miguel Andres-Bray, et al. 2019. Confrustion in Learning from Erroneous Examples: Does Type of Prompted Self-explanation Make a Difference? *AIED*.

34. Luis-Felipe Rodríguez and Félix Ramos. 2014. Development of Computational Models of Emotions for Autonomous Agents: A Review. *Cognitive Computation* 6, 3: 351–375.

35. Laura Sesma-Sanchez and Dan Hansen. 2018. Binocular model-based gaze estimation with a camera and a single infrared light source. 1–5.

36. Jai-Yeol Son and Sung S. Kim. 2008. Internet Users' Information Privacy-Protective Responses: A Taxonomy and a Nomological Model. *MIS Quarterly* 32, 3: 503–529.

37. A. S. Tolba, A. H. El-baz, and A. A. El-harby. 2006. *Face Recognition: A Literature Review*. .

38. Patrick Wied. Dynamic Heatmaps for the Web. Retrieved July 15, 2020 from https://www.patrick-wied.at/static/heatmapjs/.

39. Diyi Yang, Robert Kraut, and Carolyn Rose. 2016. Exploring the Effect of Student Confusion in Massive Open Online Courses. *JEDM | Journal of Educational Data Mining* 8, 1: 52–83.

confusion. (n.d). *The American Heritage® Medical Dictionary*. (2007). Retrieved May 19 2020 from https://medical-dictionary.thefreedictionary.com/confusion on 20 May 2020

41. Emotion Detection. Retrieved June 14, 2020 from https://www.gartner.com/en/information-technology/glossary/emotion-detection.

APPENDIX – HIGH-LEVEL SOLUTION REQUIREMENTS

The ideal solution, that solves the problem statement, will adhere to the requirements set out in Table 6:

	Solution Requirement	Description
1	**Inexpensive**	The sensor is inexpensive. It must be easily acquired or, ideally, already available to the student.
2	**Inobtrusive**	The sensor is inobtrusive. Ideally the sensor does not impact or interrupt the normal learning process of the student.
3	**Scalable**	Deployment of the sensor and detection method is scalable. It can be deployed with relative ease outside of laboratory experiment environments.
4	**Content**	The detection method is independent of subject content (audio, video, images, text) .
5	**Environment**	The detection method is independent of the teaching environment (online, offline, classroom, laboratory).
6	**Demography**	The detection method is independent of demography (age, level of education, cultural background).
7	**Reliable**	The detection method is reliable and accurate and can clearly identify (1) when the confusion occurs, (2) what is being observed that is confusing, and (3) why the student is confused.

Table 6 - High-level Solution Requirements

APPENDIX – SOLUTION REQUIREMENT MEASURING SCALE

The measuring scale used during the research is shown in Table 7.

	Solution Requirement	Legend	Description
1	**Inexpensive**	○	The sensor is expensive and only used where there is serious commercial endeavor or large-scale research studies
		◔	The sensor must be purchased separately and is affordable to small business or research studies
		◑	The sensor must be purchased separately and is affordable for most consumers
		◕	There is very limited cost associated with the sensor, or the sensor is already available at the sunken cost of acquiring a different device
		●	There is zero cost associated with obtaining the sensor
2	**Inobtrusive**	○	The end user is completely aware of the sensor and is physically wired to it
		◔	The end user is completely aware of the sensor but is not physically wired to it
		◑	The end user is unaware of the sensor, but some initial setup is required to allow for the sensor to function correctly
		◕	The end user is unaware of the sensor but must perform an action that interrupts or impacts what would have been a natural course of action
		●	The end user is completely unaware of the testing and it does not impact or interfere with their normal course of action

3	**Scalable**	○	The solution is deployed on a case base case basis and requires careful setup and calibration each time per test subject
		◔	The solution is deployed on a case base case basis and requires careful setup and calibration each time per environment, but suites multiple test subjects
		◑	The solution requires some relatively easy initial installation, but needs to be done on a test subject per test subject basis
		◕	The solution requires some relatively easy initial installation, and is done only once for all test subjects
		●	The solution can be deployed to millions of users without the installation of special software or hardware
4	**Content**	○	The solution works only for one of the mediums or stimuli (audio, video, images, text)
		◔	The solution works only for two of the mediums or stimuli (audio, video, images, text)
		◑	The solution works only for three of the mediums or stimuli (audio, video, images, text)
		◕	The solution can work irrespective of medium or stimulus (audio, video, images, text), but not simultaneously for more than one medium
		●	The solution can work irrespective of medium or stimulus (audio, video, images, text), and works for any simultaneous combination of mediums
5	**Environment**	○	Highly configured laboratory environment with high administrative requirements
		◔	Classroom environment which requires some administration
		◑	Offline or online environment with some or little administrative requirement
		◕	Online environment which requires some administration or collation of results to establish a pattern
		●	Open and online environment whit no special administration requirements
6	**Demography**	○	The sensor is limited significantly by one or more demographic aspects of the test subject
		◔	The sensor and detection method must be configured based on a high number of demographic aspects of each test subject
		◑	The sensor and detection method has some dependence on the demographic of the test subject, and outcomes must be reviewed post testing
		◕	The sensor and detection method has some dependence on the demographic of the test subject, but outcomes do not have to be reviewed post testing
		●	The sensor and detection method is completely independent of the demographic of the test subject
7	**Reliable**	○	The detection method lacks proof of reliability
		◔	The detection method is proved to be somewhat reliable and accurate and can identify when the confusion occurs but not what is being observed at that time

◑	The detection method is proved to be somewhat reliable and accurate and can identify when the confusion occurs, and what is being observed at that time
◕	The detection method is reliable and accurate and can clearly identify (1) when the confusion occurs, (2) what is being observed that is confusing, but not why the student is confused. The measurement is subconscious and cannot be influenced by external factors.
●	The detection method is reliable and accurate and can clearly identify (1) when the confusion occurs, (2) what is being observed that is confusing, and (3) why the student is confused. The measurement is subconscious and cannot be influenced by external factors.

Table 7 - Detailed requirement measuring scale

APPENDIX – SOLUTION REQUIREMENT WEIGHTING

To ensure that reliability is a key feature of the various sensor and confusion detection methods, a weighting was applied. The weight of reliability was set to be the sum of all other weights, as can be seen in Table 8.

	Solution Requirement	Order of Importance	Weight
1	Inexpensive	2	1
2	Inobtrusive	2	1
3	Scalable	2	1
4	Content	2	1
5	Environment	2	1
6	Demography	2	1
7	Reliable	1	2
	Total Weight		**8**

Table 8- Solution Requirement weights

An ordinal scale was used to assign a value, where the lowest possible value was assigned zero (0), and the highest possible value assigned was four (4). Multiplying this value with the weight, the weight average was calculated for each sensor and detection method as shown in Table 9.

Sensor	Detection Method	Inexpensive	Inobtrusive	Scalable	Content	Environment	Demography	Reliable	Weighted Average
EMG	Electrical skeletal muscle activity	0	0	1	4	0	4	6	1.9
EEG	Electrical brain activity	2	1	1	4	2	4	6	2.5
ECG	Electrical heart activity	0	0	1	4	0	4	6	1.9
Self-Report	Muddy cards	4	3	4	3	3	2	4	2.9
	Clicking behavior	4	4	4	2	3	2	4	2.9
	Sentiment Analysis	4	4	4	1	3	2	4	2.8

Sensor	Detection Method	Inexpensive	Inobtrusive	Scalable	Content	Environment	Demography	Reliable	Weighted Average
	Choice selection	4	3	4	2	4	2	4	2.9
FER	2D without gaze estimation	3	2	4	4	4	3	2	2.9
	3D without gaze estimation	2	4	4	4	4	3	4	3.1
	2D with gaze estimation	3	2	4	4	4	3	4	3.1
	3D with gaze estimation	2	4	4	4	4	3	4	3.1

Table 9 - Weighted average fit for each sensor and detection method

APPENDIX – ARCHITECTURE

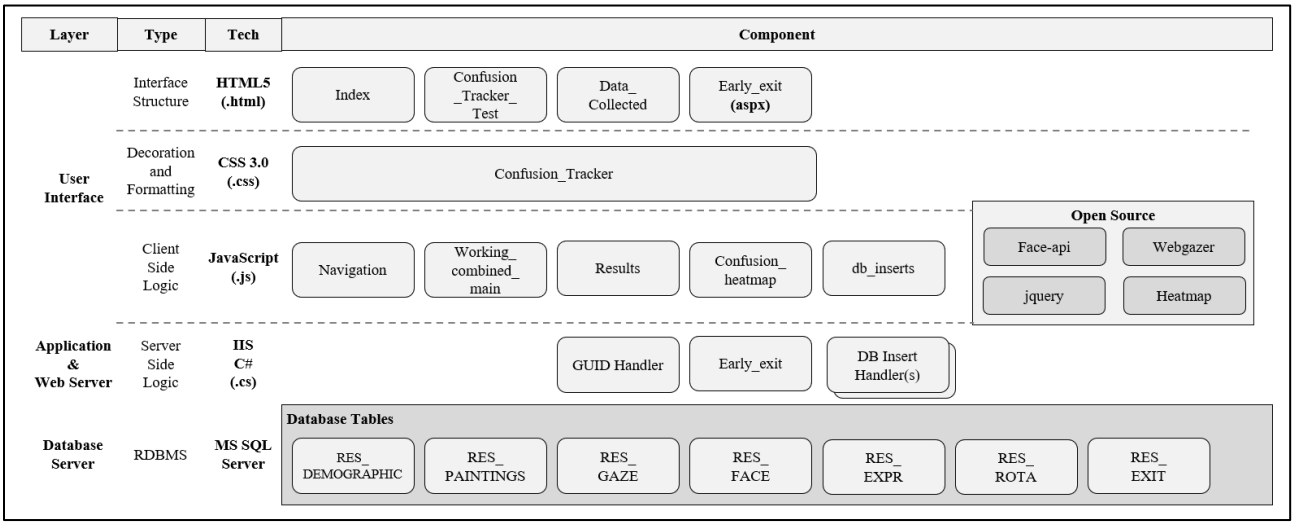


Figure 6 - High-level Architecture and Software components

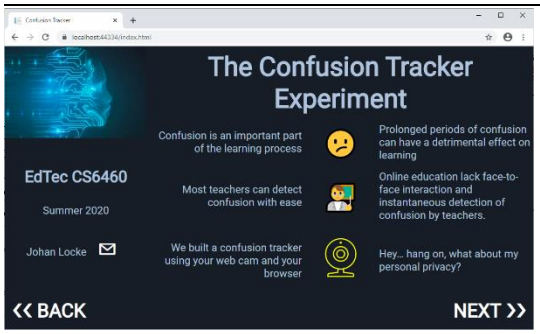
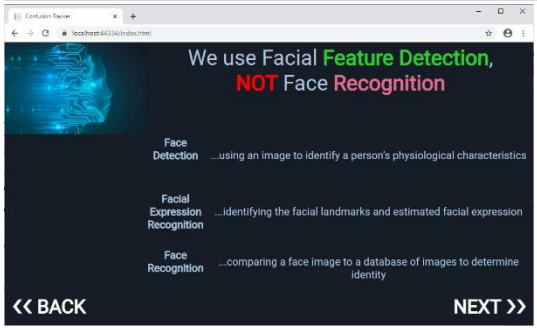
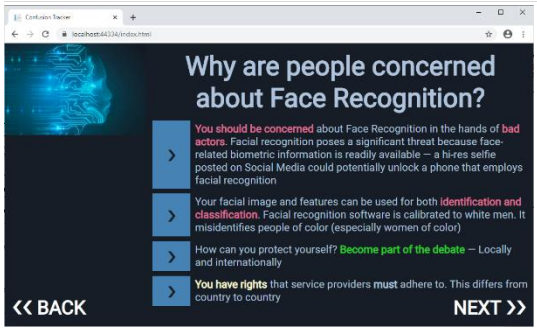
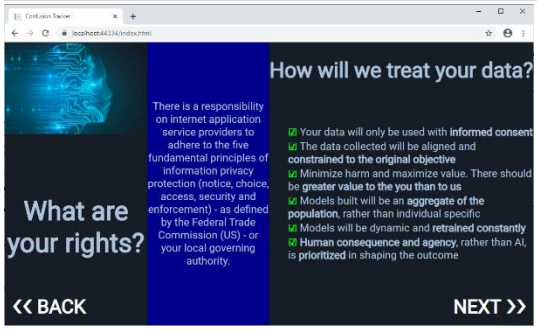
User Interface Layer

The system has been developed to be 100% web browser-based without the need for the user to install any additional components. The user interface has been developed using HTML5, CSS 3.0 and JavaScript.

The user interface has been split into two main sections. The first section deals with an introduction into concept of confusion and privacy concerns the stem from the use of facial recognition technology. The second section runs the confusion tracker experiment itself.

Index.html

This is the landing page of the website and works like a slide show presentation. Users can used forward and back button to navigate through the content.

Slide	Description
	Landing slide to explain to users the concept of confusion.
	Slide to explain the difference between "Face Detection", "Facial Expression Recognition" and "Face Recognition"
	Slide to explain why having concerns about Face Recognition are substantiated.
	Slide to explain what a end-user's rights are when it comes to information privacy, as well as a commitment and principles to how we will treat their data

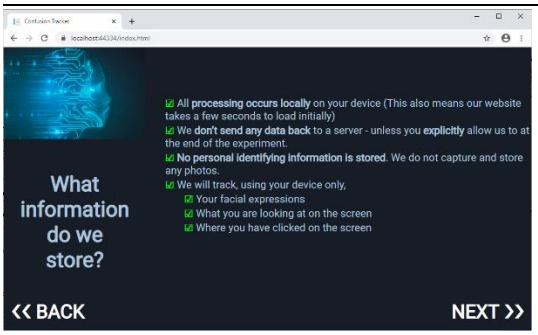
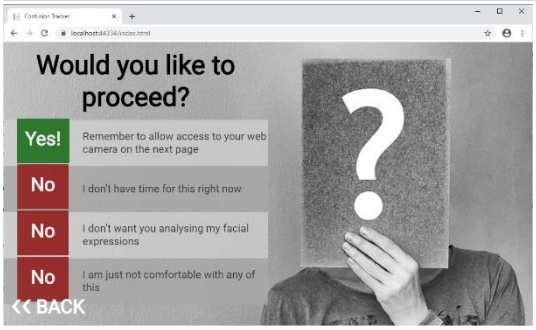
Slide	Description
	Slide to explain to users what information will be captured during the experiment, and what information will be stored
	A final slide to give the user the option to firmly agree to continuing with the experiment or exit on their own terms. For statistical reasons we capture three possible exit options: • I don't have time for this right now • I don't want you analyzing my facial expressions • I am just not comfortable with any of this

Table 10: Slide Components of the Landing Page - index.html

Early_exit.aspx

The page is a simple C# web form which captures (1) the reasons end-users choose not to continue with the experiment, or if they completed the experiment, (2) the fact that they have exited without wanting their experiment data stored.

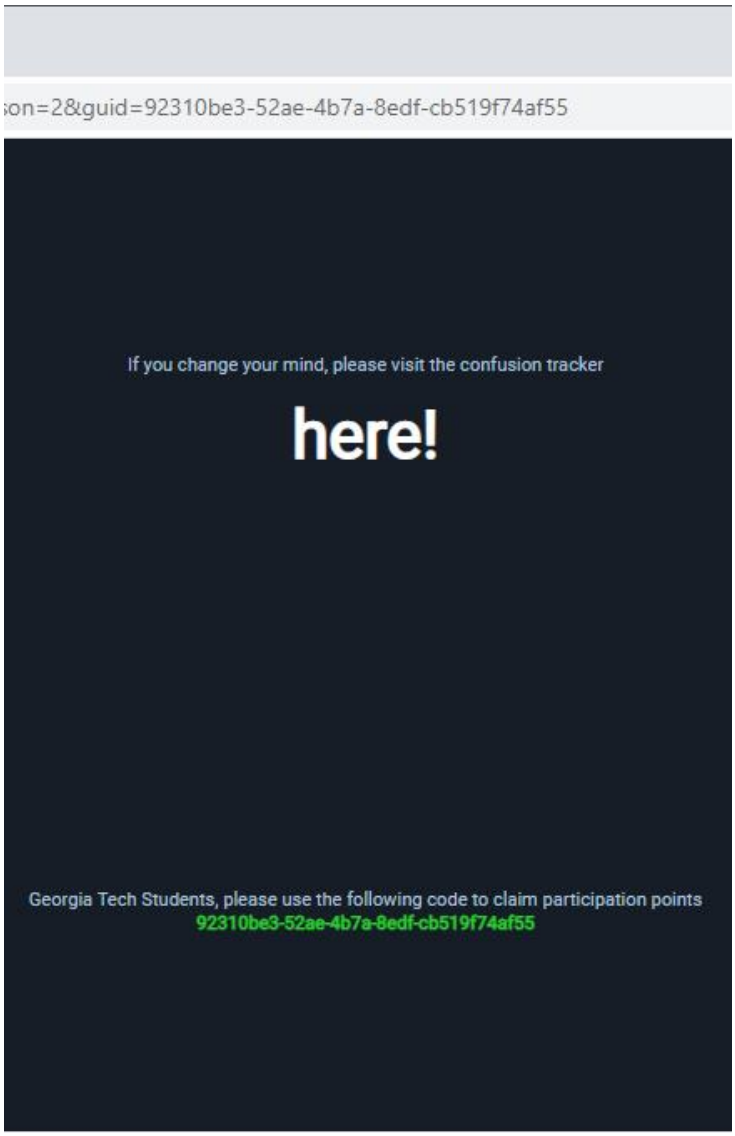


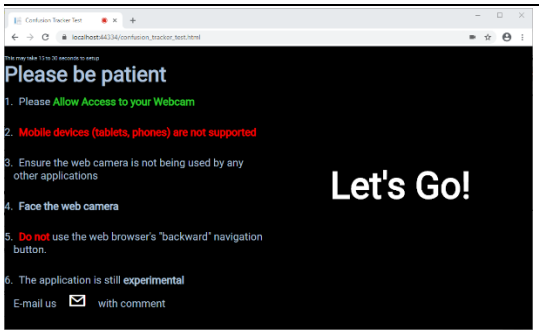
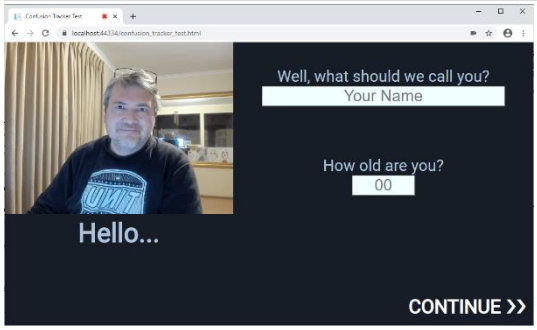
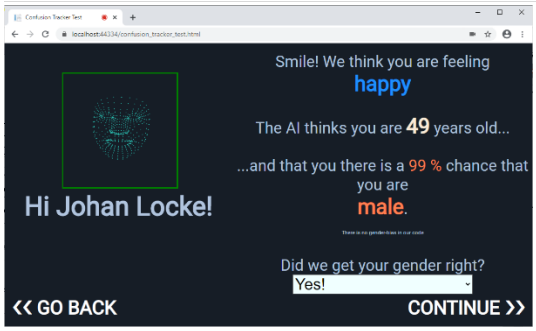
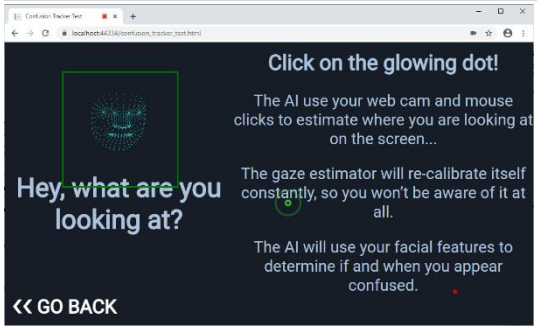
Figure 7 - Screenshot of user's Early Exit decision captured

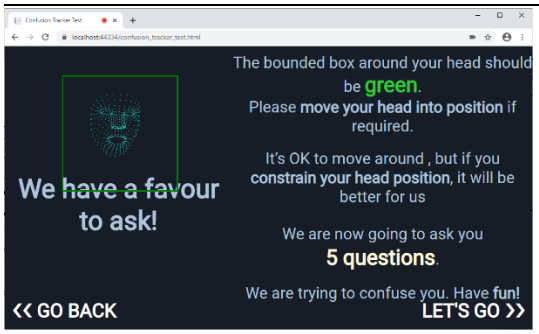
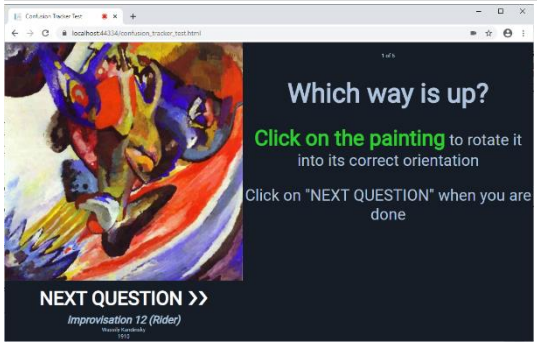
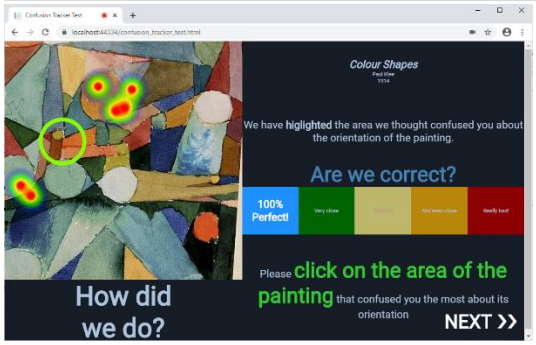
To facilitate the rewarding of Georgia Institute of Technology students for participating in the study, a unique code is generated.

Confusion_tracker_test.html

This single HTML page is the heart of the experiment. It utilises four (4) open source JavaScript libraries (as described in "Open source JavaScript Libraries"), without modification.

A single web page architecture was chosen to retain gaze calibration data as the user moves through the various process stages.

Slide	Description
	An initial loading slide of the experiment during which: • All necessary client-side JavaScript files and models are downloaded • The camera interface, face tracking and gaze estimation is initialized
	A slide to capture basic demographic information in a user-friendly way.
	A slide to start giving the user a view of the information the system is capturing, as well as visual confirmation that the system is working
	A slide to calibrate the gaze estimation software. The experiment uses a 500x500 pixel area to display images (to test confusion). To ensure that gaze is tracked with higher accuracy in this region, dots are placed in this square are on this screen which the user needs to click on.

Slide	Description
	As the open source component utilized to track gaze is not highly tolerant of pose variation, this slide requests users to constrain their head position as much as possible.
	This slide is the heart of experiment capture. As explained in "Methodology", it displays a series of five (5) abstract paintings which the user needs to orientate into the artists original intended orientation.
	This slide displays a heatmap (refer to "Client-side Logic") using the estimated areas of confusion, calculated as described in "Estimating the area of confusion". It allows users to rate the accuracy of the estimation, and self-report the area that confused them the most. The slide is shown for each of the paintings and users cannot proceed until a selection is made.

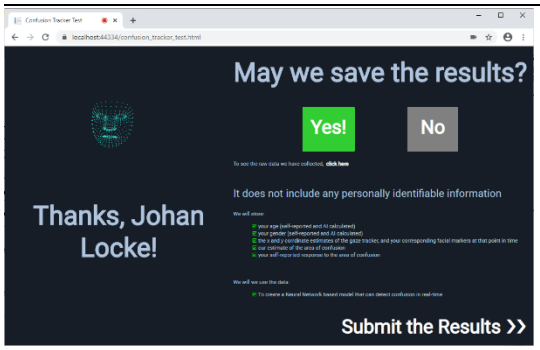
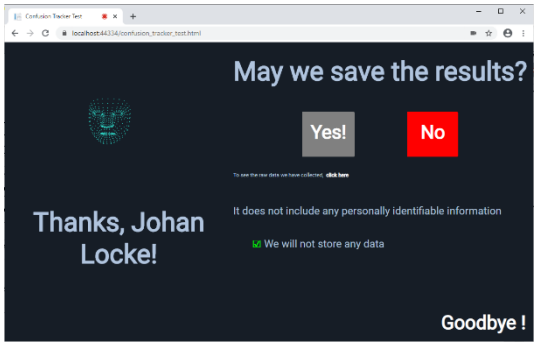
Slide	Description
 	To meet the objective of ensuring personal privacy, users are given a final chance to choose if they want to submit their collected data. If "yes" is selected, a detailed explanation is given again as to what data will be stored, and what it will be used for in further studies. If "no" is selected, it is confirmed back to the user that no data will be stored and on their exit is recorded using "Early_exit.aspx" Users can under either option selected to view the raw data collected, as described in "Data_collected.html".

Table 11 - Slide Components of the Experiment - Confusion_tracker_test.html

Data_collected.html

The page presents users with the actual raw data captured during the experiment. It retrieves the data from the client browser and never stores it to permanent storage.

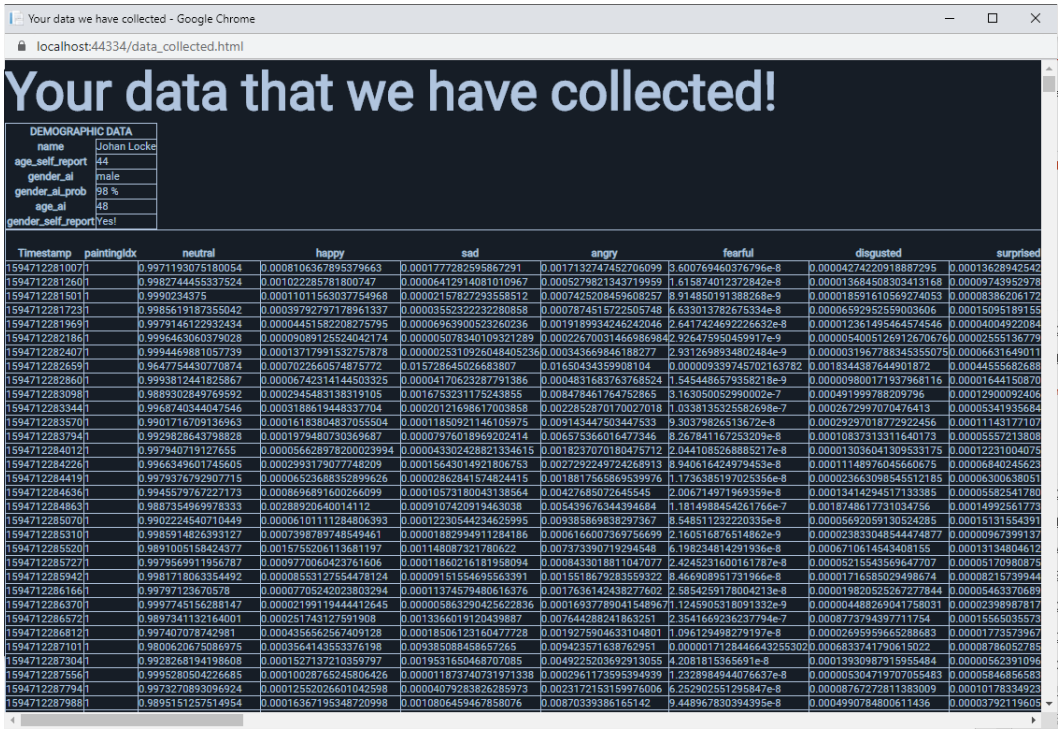


Figure 8 - Screenshot of data collected during experiment for the active participant.

Client-side Logic

All process logic has been either newly developed JavaScript page or the re-use of existing open source libraries

Open source JavaScript Libraries

Library	Purpose
jQuery [18]	We used jQuery because of its ease of use to make Ajax asynchronous calls to C# handlers to store data
webgazer.js [30]	We used webgazer.js to provide gaze estimations as well as a visual presentation of the user's facial features, as can be seen in Figure 9.
face-api.js [25]	This library provided demographic information (age, gender), facial features and facial expression estimation. Of the facial features, we only permanently stored the left eye, right eye, left eyebrow, and right eyebrow, if the user permitted the storing of the experiment data
heatmap.js [38]	This library provided a heatmap image that show the estimated area of confusion, calculated as described in "**Error! Reference source not found.**".

Figure 9 - Screenshot of Facial Feature Image produced by webgazer.js

Table 12 - Open source JavaScript libraries used in the development of the system

Custom developed JavaScript scripts

Script	Purpose
navigation.js	This custom develop script handles all display related logic such as navigating between slides, hiding/unhiding objects. Base slide navigation is based off a W3schools.com tutorial located at https://www.w3schools.com/howto/howto_js_slideshow.asp

Script	Purpose
working_combined_main.js	This script initializes the face-api and webgazer.js libraries and creates event handlers that estimate the current values for gaze location, facial features, facial expression, age and gender. Base webgazer.js functionality is based off the specific example implementation given by Papoutsaki et al (2016), located at https://webgazer.cs.brown.edu/calibration.html Base face-api functionality was built on top of the implementation located at https://dev.to/karkranikhil/face-recognition-using-javascript-33n5 [19]
results.js	Stores all experiment results (Demographics, painting self-report confusion location, gaze estimations, facial features, facial expressions, rotation events) This script also calculates the estimated area of confusion, using the methodology described in "**Error! Reference source not found.**".
confusion_heatmap.js	Displays a heatmap based off the estimated area of confusion data using the open source library heatmap.js. Code is based of the example given by Wied (n.d.), located at https://www.patrick-wied.at/static/heatmapjs/examples.html
db_inserts.js	The script manages AJAX (Asynchronous JavaScript And XML) calls to C# (programming language) handlers for insert data into the database, as well as generating a unique GUID for each session.

Table 13 - Custom developed JavaScript scripts used in the development of the system

Web & Application server layer

Microsoft's C# was used in conjunction with Microsoft IIS to develop the application server logic. The solution was deployed to an Azure (cloud-based) environment.

Component	Purpose
Guid_handler.ashx	Returns a unique GUID (Global Unique Identifier) when called
early_Exit.aspx.cs	Stores the reason a user elects not to continue with normal processing
DB_insert_expr_handler.ashx	Accepts facial expression data recorded during the experiment, in JSON (JavaScript Object Notation) format, and stores it to the RES_EXPR table.
DB_insert_face_handler.ashx	Accepts facial features data recorded during the experiment, in JSON format, and stores it to the RES_FACE table.
DB_insert_gaze_handler.ashx	Accepts gaze estimation data recorded during the experiment, in JSON format, and stores it to the RES_GAZE table.
DB_insert_handler.ashx	Accepts demographic data recorded during the experiment, in JSON format, and stores it to the RES_DEMOGRAPHIC table.
DB_insert_paintings_handler.ashx	Accepts painting source, and user self-reported confusion location data recorded during the experiment, in JSON format, and stores it to the RES_PAINTINGS table.

Component	Purpose
DB_insert_rota_handler.ashx	Accepts painting rotation event data recorded during the experiment, in JSON format, and stores it to the RES_ROTA table.

Table 14 - Custom developed Application Server Components

Database layer

A Microsoft SQL Server database was used, hosted on Azure. The database contains seven (7) different database tables and one (1) view.

- All tables store a GUID at row-level to uniquely identify and link the data between the various tables.
- All table entries store a timestamp (in UTC-time format) to record the time at which the measurment was taken.

Database Object Name	Type	Purpose
RES_DEMOGRAPHIC	Table	Contains self-reported demograpic information of the respondent, as well as AI (Artificial Intellegence estimated) values for age and gender.
RES_EXIT	Table	Contains the reason for exiting the experiment early by the respondent

Code	Reason
0	I don’t have time for this right now
1	I don’t want you analyzing my facial expressions
2	I am just not comfortable with any of this
3	Respondent elects not to save collected data post the experiment

RES_EXPR	Table	Contains facial expression estimate data, over time
RES_FACE	Table	Contains facial feature location data, over time
RES_GAZE	Table	Contains gaze estimation data, over time
RES_PAINTINGS	Table	Contains data about each painting displayed to the respondent as well as the self-reported location of confusion, and their view of how accurate the estimated area of confusion for a specific painting was
RES_ROTA	Table	Contains painting rotation event data, for each painting, over time
RESPONSES	View	Create a single view of all respondents. Code 0,1,2,3 correlates with the RES_EXIT table. Code 5 indicates the respondent exited the experiment by saving their data to the database.

Table 15 - Database tables used to capture experiment responses

APPENDIX – IMAGES OF ABSTRACT PAINTINGS USED DURING EXPERIMENTATION

All artworks fall with the “Public Domain”, as classified by wikiart.com:

WikiArt.org allows unlimited copying, distributing and displaying of the images of public domain artworks. Artworks protected by copyright are supposed to be used only for contemplation. Images of that type of artworks are prohibited for copying, printing, or any kind of reproducing and communicating to public since these activities may be considered copyright infringement.

Artist	Year	Painting Title	Cropped Image used during Experiment
Wassily Kandinsky	1910	Improvisation 12 (Rider)	
		Original: https://www.wikiart.org/en/wassily-kandinsky/improvisation-12-rider-1910	
Paul Klee	1914	Colour Shapes	
		Original: https://www.wikiart.org/en/paul-klee/colour-shapes-1914	

Artist	Year	Painting Title	Cropped Image used during Experiment
Hans Hofmann	1942	Provincetown	
		Original: https://www.wikiart.org/en/hans-hofmann/provincetown-1942	
Francis Picabia	1913	Ballerina on an Ocean Liner	
		Original: https://www.wikiart.org/en/francis-picabia/ballerina-on-an-ocean-liner-1913	

Artist	Year	Painting Title	Cropped Image used during Experiment
Paul Klee	1919	Swamp Legend	

https://www.wikiart.org/en/paul-klee/swamp-legend-1919

Table 16 - Artwork used during testing stage of the experiment

APPENDIX – TEST POPULATION

Group	Demographic	Contact method	Estimated number invited to participate
Students enrolled in the CS-6460 course	Post-graduate students Age 21-60	Piazza post	300
1st Degree LinkedIn contacts of Johan Locke	Professional Age 18-75	LinkedIn article	5700
Friends and Family of Johan Locke	South African and Australian Age 9-75	Email and face-to-face contact	40

Table 17 - Test Population Groups

APPENDIX – RANDOM LOCATION GENERATOR

Python Script

```
import numpy as np
import random
iter  = 100000 # One hundred thousand
max_x = 500
max_y = 500
data  = np.zeros((iter,5),float)
```

```
for i in range(iter):
    data[i,:4] = [random.randint(0, max_x),
                  random.randint(0, max_y),
                  random.randint(0, max_x),
                  random.randint(0, max_y)]

data[:,4] = ((data[:,0] - data[:,2])**2 + (data[:,1] - data[:,3])**2 )**(1/2)

print("Mean x:          {}".format(data[:,0].mean()))
print("StdDev x:        {}".format(data[:,0].std()))
print("Mean y:          {}".format(data[:,1].mean()))
print("StdDev y:        {}".format(data[:,1].std()))
print("Mean distance:   {}".format(data[:,4].mean()))
print("StdDev distance: {}".format(data[:,4].std()))
```

Output

```
Mean x:          250.190644
StdDev x:        144.67068796015752
Mean y:          249.748038
StdDev y:        144.69942535183253
Mean distance:   261.29953593133257
StdDev distance: 124.20633665022983
```

APPENDIX – EXTERNAL STORED DATA AND SCRIPTS

Resource Type		Location and further information
Application (Public Website)		https://confusiontrackerapplication20200706161558.azurewebsites.net/ or https://www.confusion-tracker.com/
Database Server		MS SQL Server (Azure) cfserverjohanlocke.database.windows.net Database: cfdb
Source Material		https://1drv.ms/u/s!ArH1csDkvUdWhu19UByJY29i4MmSMQ?e=s3lL5X
	Database Scripts	Scripts to create database tables and read statistics "/09 Final Project/sourcecode/WebApplication1/database table scripts"
	Application Source Code	"/09 Final Project/sourcecode/WebApplication1/"
	peerSurvey Data	"/08 Final Paper/peerSurvey data collected/peersurvey.csv" "/08 Final Paper/peerSurvey data collected/survey data/xlsx"
	Final Paper	"/08 Final Paper/jlocke33 Final Paper SIGHI format.pdf"
	Final Presentation	"/10 Final Presentation/jlocke33 Final presentation.pdf" "/10 Final Presentation/jlocke33 Final presentation.mp4"
	Data Collected	"/09 Final Project/raw_data/data_dump.xlsx"